



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2016

---

## **Data-collection strategy for challenging native SAD phasing**

Olieric, Vincent ; Weinert, Tobias ; Finke, Aaron D ; Anders, Carolin ; Li, Dianfan ; Olieric, Natacha ; Borca, Camelia N ; Steinmetz, Michel O ; Caffrey, Martin ; Jinek, Martin ; Wang, Meitian

**Abstract:** Recent improvements in data-collection strategies have pushed the limits of native SAD (single-wavelength anomalous diffraction) phasing, a method that uses the weak anomalous signal of light elements naturally present in macromolecules. These involve the merging of multiple data sets from either multiple crystals or from a single crystal collected in multiple orientations at a low X-ray dose. Both approaches yield data of high multiplicity while minimizing radiation damage and systematic error, thus ensuring accurate measurements of the anomalous differences. Here, the combined use of these two strategies is described to solve cases of native SAD phasing that were particular challenges: the integral membrane diacylglycerol kinase (DgkA) with a low Bijvoet ratio of 1% and the large 200 kDa complex of the CRISPR-associated endonuclease (Cas9) bound to guide RNA and target DNA crystallized in the low-symmetry space group C2. The optimal native SAD data-collection strategy based on systematic measurements performed on the 266 kDa multiprotein/multiligand tubulin complex is discussed.

DOI: <https://doi.org/10.1107/S2059798315024110>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-125047>

Journal Article

Published Version

Originally published at:

Olieric, Vincent; Weinert, Tobias; Finke, Aaron D; Anders, Carolin; Li, Dianfan; Olieric, Natacha; Borca, Camelia N; Steinmetz, Michel O; Caffrey, Martin; Jinek, Martin; Wang, Meitian (2016). Data-collection strategy for challenging native SAD phasing. *Acta Crystallographica. Section D: Structural Biology*, 72(Pt 3):421-429.

DOI: <https://doi.org/10.1107/S2059798315024110>

# Data-collection strategy for challenging native SAD phasing

Vincent Olieric,<sup>a,\*</sup> Tobias Weinert,<sup>a,†</sup> Aaron D. Finke,<sup>a</sup> Carolin Anders,<sup>b</sup> Dianfan Li,<sup>c</sup> Natacha Olieric,<sup>d</sup> Camelia N. Borca,<sup>a</sup> Michel O. Steinmetz,<sup>d</sup> Martin Caffrey,<sup>c</sup> Martin Jinek<sup>b</sup> and Meitian Wang<sup>a</sup>

Received 12 May 2015

Accepted 15 December 2015

<sup>†</sup> These authors contributed equally to this work.

<sup>§</sup> Present address: National Center for Protein Sciences, State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, People's Republic of China.

**Keywords:** native SAD phasing; anomalous signal; data-collection strategy; complex; membrane protein.

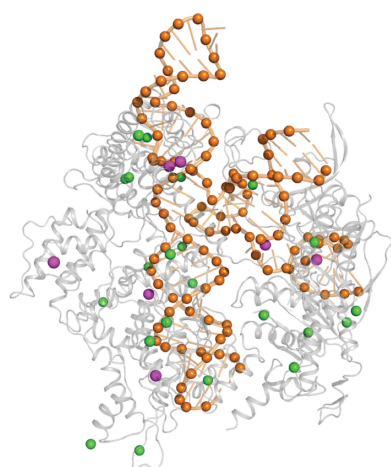
**Supporting information:** this article has supporting information at journals.iucr.org/d

<sup>a</sup>Swiss Light Source, Paul Scherrer Institut, Villigen PSI, Switzerland, <sup>b</sup>Department of Biochemistry, University of Zurich, Zurich, Switzerland, <sup>c</sup>Membrane Structural and Functional Group, Schools of Medicine and Biochemistry and Immunology, Trinity College, Dublin, Ireland, and <sup>d</sup>Laboratory of Biomolecular Research, Department of Biology and Chemistry, Paul Scherrer Institut, Villigen PSI, Switzerland. \*Correspondence e-mail: vincent.olieric@psi.ch

Recent improvements in data-collection strategies have pushed the limits of native SAD (single-wavelength anomalous diffraction) phasing, a method that uses the weak anomalous signal of light elements naturally present in macromolecules. These involve the merging of multiple data sets from either multiple crystals or from a single crystal collected in multiple orientations at a low X-ray dose. Both approaches yield data of high multiplicity while minimizing radiation damage and systematic error, thus ensuring accurate measurements of the anomalous differences. Here, the combined use of these two strategies is described to solve cases of native SAD phasing that were particular challenges: the integral membrane diacylglycerol kinase (DgkA) with a low Bijvoet ratio of 1% and the large 200 kDa complex of the CRISPR-associated endonuclease (Cas9) bound to guide RNA and target DNA crystallized in the low-symmetry space group C2. The optimal native SAD data-collection strategy based on systematic measurements performed on the 266 kDa multiprotein/multiligand tubulin complex is discussed.

## 1. Introduction

Native SAD (single-wavelength anomalous diffraction) is a *de novo* macromolecular structure-determination method in which the phase problem is solved by exploiting the anomalous diffraction signal of light atoms present in the crystal (for a review of anomalous diffraction, see Hendrickson, 2014). Unlike other *de novo* methods, structure solution by native SAD does not require the incorporation of exogenous heavy atoms, which can be tedious and can lead to non-isomorphism. However, native SAD phasing has its own challenges. Firstly, the absorption edge, where anomalous scattering is maximized, is below 2.5 keV (>5 Å) for elements such as sulfur and phosphorus, which is beyond the reach of most current macromolecular crystallography (MX) synchrotron beamlines. Data collection is also hindered by sample and air absorption of X-rays at such low energies. Thus, native SAD data collection is usually performed at 'compromise' energies of around 6 keV (2.066 Å; Mueller-Dieckmann *et al.*, 2005; Liu *et al.*, 2014; Weinert *et al.*, 2015) to maximize the anomalous signal while minimizing absorption. Alternatively, home sources, which typically operate with Cu K $\alpha$  (1.54 Å) or Cr K $\alpha$  (2.29 Å) radiation, can also be used for such measurements. Secondly, because data collection is performed far from the absorption edges of these light elements, the resulting anomalous signal is small and accurate measurements of the reflection intensities and their differences are



OPEN ACCESS

essential. To obtain this level of accuracy, random and systematic errors must be minimized. However, typical data-collection protocols at third-generation synchrotrons, usually performed around a single axis, yield anomalous data that can be adversely affected by radiation damage if insufficient care is given to the X-ray dose delivered to the crystal. Consequently, successful cases of phasing by native SAD remain rare; currently, about 150 native SAD structures have been deposited in the Protein Data Bank (Berman *et al.*, 2000).

Recent advances in data-collection strategies have breathed new life into native SAD phasing (reviewed by Rose *et al.*, 2015). The multi-crystal averaging approach involves merging data sets from statistically equivalent crystals (Liu *et al.*, 2012, 2013), enhancing the anomalous signal-to-noise ratio while minimizing anomalous signal decay owing to radiation damage. The key to successful multi-crystal averaging having enough isomorphous crystals available for data collection. This method is possible because data-merging and data-processing algorithms are robust enough to accommodate a large number of distinct crystal data sets. In particularly challenging cases, data sets from as many as 18 (Akey *et al.*, 2014) and 32 (El Omari *et al.*, 2014) separate crystals were required to obtain sufficient anomalous signal for successful phasing. An extreme case is the recent native SAD phasing of lysozyme using serial crystallography data collected at room temperature, where 2° of data obtained individually from 992 randomly oriented crystals with an average size of 20 µm were merged (Huang *et al.*, 2015). A second approach involves collecting many low-dose data sets (<0.5 MGy per 360°) from a single crystal in multiple orientations, mitigating radiation damage while also reducing systematic errors in data collection (Debreczeni *et al.*, 2003; Brockhauser *et al.*, 2013; Weinert *et al.*, 2015; Finke *et al.*, 2016). By varying the crystal orientation, it is possible to measure the same reflections in different diffraction geometries on different areas of the detector and with different sample absorption, thereby reducing systematic error. This approach has proven to be successful for 11 real-life cases, including a 266 kDa multi-protein/multiligand complex, the largest structure solved by native SAD to date (Weinert *et al.*, 2015). The method benefited from instrumentation developments such as a high-precision PRiGo multi-axis goniometer (Waltersperger *et al.*, 2015) and a photon-counting, noise-free pixel-array detector calibrated for low energies (Henrich *et al.*, 2009), both of which were used in the current study.

Despite the many recent improvements in data-collection strategies, difficult cases such as large anomalous scattering substructures, low-symmetry space groups, low diffraction resolution, small crystals, low Bijvoet ratios or some combination thereof still remain a challenge for native SAD phasing. Fortunately, it is possible to combine the two strategies introduced above to extract as much anomalous signal as possible with the highest possible accuracy from as many crystals as needed for successful phasing (Klinke *et al.*, 2015). Here, we present two challenging native SAD cases solved by collecting low-dose data at multiple orientations from three crystals each: the integral membrane diacylglycerol kinase

DgkA (Li *et al.*, 2013;  $2 \times 42$  kDa asymmetric unit, space group  $P2_12_12_1$ , 2.6 Å resolution) and the large Cas9–RNA–DNA complex (Anders *et al.*, 2014; 200 kDa asymmetric unit, space group  $C2$ , 2.2 Å resolution). In addition, we demonstrate the benefits of these combined low-dose multi-orientation, multi-crystal data-collection strategies using systematic measurements performed on the multiprotein/multiligand tubulin–RB3–tubulin tyrosine ligase complex T<sub>2</sub>R–TTL (Prota, Bargsten *et al.*, 2013; Weinert *et al.*, 2015; 266 kDa asymmetric unit, space group  $P2_12_12_1$ , 2.3 Å resolution).

## 2. Materials and methods

### 2.1. Crystallographic data collection, data processing and analysis

All experiments were performed on the super-bending-magnet beamline X06DA (PXIII) at the Swiss Light Source (SLS), Villigen PSI, Switzerland operating at 2.4 GeV with 400 mA top-up mode. The beamline has a double channel-cut Si(111) monochromator with an energy resolution of  $1.4 \times 10^{-4}$  and an elliptical beam of  $90 \times 50$  µm (FWHM). Data were collected at 100 K with a wavelength of 2.066 Å (6 keV) and a flux of  $\sim 1.5 \times 10^{10}$  photons s<sup>-1</sup>. Multi-orientation data collection was carried out at various  $\chi$  and  $\phi$  settings of the multi-axis PRiGo goniometer (Waltersperger *et al.*, 2015) and on a PILATUS 2M-F detector (Henrich *et al.*, 2009) operated in shutterless mode at a frame rate of 10 or 20 Hz and at a sample-to-detector distance of 120 mm. The data were processed using *XDS* and scaled and merged with *XSCALE* (Kabsch, 2010). The high-resolution data cutoff was based on the statistical indicators  $CC_{1/2}$  and  $CC^*$  (Karplus & Diederichs, 2012). Substructure determination and phasing were performed with *SHELXC/D/E* (Sheldrick, 2010) using the *HKL2MAP* interface (Pape & Schneider, 2004). The anomalous peak heights were calculated using *AnoDe* without a resolution cutoff (Thorn & Sheldrick, 2011). The number of correct sites found with *SHELXD* was assessed using *SITCOM* (Dall'Antonia & Schneider, 2006). Refinement and model map cross-correlation calculations were performed using *PHENIX* (Adams *et al.*, 2010). X-ray dose was estimated with *RADDOSE-3D* (Zeldin *et al.*, 2013). The Bijvoet ratio was estimated using the formula

$$(|\Delta F^{\pm}|)/(|F|) = (2N_A/N_P)^{1/2} \times (f''/Z_{\text{eff}}), \quad (1)$$

where  $f''$  is the imaginary scattering contribution of sulfur (0.95 e at 6 keV),  $N_A$  is the number of anomalous scattering atoms,  $N_P$  is the number of non-H atoms in the asymmetric unit, and  $Z_{\text{eff}}$  is the effective number of electrons of the average protein atom (6.7; Hendrickson & Teeter, 1981). Structure figures were prepared and rendered using *PyMOL* (DeLano, 2002).

**2.1.1. DgkA.**  $P2_12_12_1$  crystals of a DgkA mutant (A41C, C46A, I53V, I70L, M96L, V107D, C113A) measuring  $\sim 100 \times 200 \times 50 \times 20$  µm (Supplementary Fig. S1a) were grown by the *in meso* or lipid cubic phase method as described elsewhere (Li *et al.*, 2013), harvested and snap-cooled in liquid

Table 1

Summary of data collection and statistics.

Values in parentheses are for the last shell.

Data set	DgkA	Cas9–RNA–DNA	T <sub>2</sub> R–TTL (high dose, low multiplicity, single orientation)†	T <sub>2</sub> R–TTL (low dose, high multiplicity, multi-orientation)†
No. of crystals	3	3	3	3
Total oscillation ( $\chi/\varphi$ orientations)	Crystal 1: 10 $\times$ 360°, 0.1 s, 0.1° ( $\chi = 0^\circ$ )  Crystal 2: 6 $\times$ 360°, 0.2 s, 0.2° ( $\chi = 0^\circ$ ), 2 $\times$ 360°, 0.2 s, 0.2° ( $a^*$ aligned)  Crystal 3: 4 $\times$ 360°, 0.1 s, 0.1° ( $\chi = 0^\circ$ ), 2 $\times$ 360°, 0.2 s, 0.2° ( $\chi = 10^\circ$ ), 2 $\times$ 720°, 0.1 s, 0.1° ( $\chi = 20^\circ$ , $\chi = 30^\circ$ )	Crystal 1 p1‡: 8 $\times$ 360°, 0.1 s, 0.2° ( $\chi = 0^\circ$ , $\chi = 10^\circ$ , $\chi = 20^\circ$ , $\chi = 30^\circ$ , $\chi = 30^\circ/\varphi = 180^\circ$ , $\chi = 20^\circ/\varphi = 180^\circ$ , $\chi = 10^\circ/\varphi = 180^\circ$ , $\chi = 5^\circ/\varphi = 180^\circ$ )  Crystal 1 p2‡: 8 $\times$ 360°, 0.1 s, 0.2° ( $\chi = 5^\circ/\varphi = 45^\circ$ , $\chi = 15^\circ/\varphi = 45^\circ$ , $\chi = 25^\circ/\varphi = 45^\circ$ , $\chi = 13^\circ$ , $\chi = 13^\circ/\varphi = 45^\circ$ , $\chi = 18^\circ$ , $\chi = 18^\circ/\varphi = 45^\circ$ , $\chi = 23^\circ/\varphi = 45^\circ$ )  Crystal 2: 8 $\times$ 360°, 0.1 s, 0.1° ( $\chi = 0^\circ$ , $\chi = 5^\circ$ , $\chi = 10^\circ$ , $\chi = 15^\circ$ , $\chi = 20^\circ$ , $\chi = 25^\circ$ , $\chi = 30^\circ$ , $\chi = 35^\circ$ )  Crystal 3: 8 $\times$ 360°, 0.1 s, 0.2° ( $\chi = 0^\circ$ , $\chi = 10^\circ$ , $\chi = 20^\circ$ , $\chi = 30^\circ$ , $\chi = 30^\circ/\varphi = 180^\circ$ , $\chi = 20^\circ/\varphi = 180^\circ$ , $\chi = 10^\circ/\varphi = 180^\circ$ , $\chi = 5^\circ/\varphi = 180^\circ$ )	Crystal 1 p1‡: 360°, 1.6 s, 0.2°  Crystal 2 p1‡: 360°, 1.6 s, 0.2°  Crystal 3 p1‡: 360°, 1.6 s, 0.2°	Crystal 1 p2‡: 8 $\times$ 720°, 0.1 s, 0.2° ( $\chi = 0^\circ$ , $\chi = 5^\circ$ , $\chi = 10^\circ$ , $\chi = 15^\circ$ , $\chi = 20^\circ$ , $\chi = 25^\circ$ , $\chi = 30^\circ$ , $\chi = 10^\circ/\varphi = 90^\circ$ )  Crystal 2 p2‡: 8 $\times$ 720°, 0.1 s, 0.2° ( $\chi = 0^\circ$ , $\chi = 10^\circ$ , $\chi = 20^\circ$ , $\chi = 30^\circ$ , $\chi = 25^\circ$ , $\chi = 15^\circ$ , $\chi = 5^\circ$ , $\chi = 15^\circ/\varphi = 90^\circ$ )  Crystal 3 p2‡: 8 $\times$ 720°, 0.1 s, 0.2° ( $\chi = 0^\circ$ , $\chi = 10^\circ$ , $\chi = 20^\circ$ , $\chi = 30^\circ$ , $\chi = 25^\circ$ , $\chi = 15^\circ$ , $\chi = 5^\circ$ , $\chi = 10^\circ/\varphi = 90^\circ$ )
Resolution (Å)	50.0–2.6 (2.67–2.60)	50–2.2 (2.30–2.20)	50–2.3 (2.39–2.30)	50–2.3 (2.39–2.30)
Space group	$P2_12_12_1$	C2	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	$a = 75.29$ , $b = 91.57$ , $c = 143.72$	$a = 177.74$ , $b = 67.57$ , $c = 188.19$ , $\beta = 111.32$	$a = 104.88$ , $b = 158.68$ , $c = 180.04$	$a = 104.86$ , $b = 158.57$ , $c = 180.31$
No. of reflections	8396768 (235461)	16157127 (388880)	4330651 (206334)	69786479 (3327104)
No. of unique reflections	29385 (2108)	205825 (24766)	255989 (27045)	258112 (28073)
Multiplicity§ (°)	142.9 (55.84)	156.7 (31.4)	33.8 (15.3)	540.7 (237.0)
Completeness (%)	99.6 (96.5)	99.3 (95.8)	99.2 (96.6)	100.0 (100.0)
$\langle I/\sigma(I) \rangle$	36.46 (4.18)	30.38 (1.88)	18.70 (1.23)	39.59 (1.19)
$R_{\text{meas}}$ (%)	22.1 (113.1)	14.1 (135.0)	11.1 (162.3)	25.5 (539.5)
$R_{\text{p.i.m.}}$ (%)	1.8 (15.1)	1.1 (24.1)	1.9 (41.5)	1.1 (35.0)
$CC_{1/2}$ (%)	100.0 (90.8)	100.0 (62.1)	99.8 (54.2)	100.0 (52.4)
$\Delta F/\sigma(\Delta F)$	1.52	1.477	1.13	2.38
$CC_{\text{ano}}$ (%)	43	44	34	63
Mosaicity (°)	0.09/0.19/0.26	0.10/0.16/0.07	0.10/0.10/0.10	0.10/0.10/0.10

† See Supplementary Data for XDS/XSCALE processing statistics of both merged and individual data sets. ‡ For crystals that were translated during measurement, the different positions are labelled p1 and p2. § Friedel pairs are counted as merged reflections.

nitrogen. Three crystals were used for data collection (Supplementary Fig. S1a). From the first crystal, which diffracted to 2.6 Å resolution, a total of ten 360°  $\omega$  scans were collected with no change in orientation. The second crystal diffracted to 2.8 Å resolution and a total of eight 360°  $\omega$  scans were collected, including two 360°  $\omega$  scans with the crystal aligned along the  $a^*$  axis in order to collect Bijvoet pairs simultaneously on the same image. Data for the third crystal, which diffracted to 2.8 Å resolution, were collected in multiple orientations: four 360°  $\omega$  scans at  $\chi = 0^\circ$  and two 360°  $\omega$  scans each at  $\chi = 10, 20$  and  $30^\circ$  (Table 1). At a data-collection speed of 1 deg s<sup>-1</sup>, the average dose per 360°  $\omega$  scan was estimated to be 0.5 MGy. Coordinates and structure factors of DgkA have been deposited in the Protein Data Bank (PDB) under accession code 5dwk.

**2.1.2. Cas9–RNA–DNA.** Crystals of a Cas9–sgRNA–target DNA complex in which the complementary (target) DNA strand was mismatched to the sgRNA guide at positions 1–3 were generated, harvested and cryoprotected as described

previously (Anders *et al.*, 2014). Three C2 crystals measuring  $\sim 200 \times 100 \times 50 \mu\text{m}$  were used for data collection (Supplementary Fig. S1b). The first, second and third crystals diffracted to 2.2, 2.4 and 2.2 Å resolution, respectively, with anomalous signal extending to  $\sim 2.9$  Å resolution. The first crystal was large enough to collect two series of multi-orientation data sets from two well separated locations on the crystal (Supplementary Fig. S1b, crystal 1). At each location, a total of eight 360°  $\omega$  scans were collected with varying  $\chi/\varphi$  settings. For the second crystal, a total of eight 360°  $\omega$  scans were collected at 5°  $\chi$  increments from 0 to 35° while keeping the  $\varphi$  orientation constant. For the third crystal, four 360°  $\omega$  scans were collected at  $\varphi = 0^\circ$  and  $\chi = 0, 10, 20$  and  $30^\circ$ , and four 360°  $\omega$  scans were collected at  $\varphi = 180^\circ$  and  $\chi = 0, 10, 20$  and  $30^\circ$  (Table 1). At a data-collection speed of 2 deg s<sup>-1</sup>, the average dose per 360°  $\omega$  scan was estimated to be 0.25 MGy. The coordinates and structure factors of Cas9–RNA–DNA have been deposited in the PDB under accession code 5fq5.



**2.1.3. T<sub>2</sub>R-TTL.** Rod-shaped  $P2_12_12_1$  T<sub>2</sub>R-TTL crystals measuring  $\sim 700 \times 100 \times 100 \mu\text{m}$  (Supplementary Fig. S1c) were grown, cryoprotected and snap-cooled in liquid nitrogen as described elsewhere (Prota, Bargsten *et al.*, 2013). Reproducible mounting and cryoprotection was critical as T<sub>2</sub>R-TTL crystals are extremely sensitive to variation of these conditions (Weinert *et al.*, 2015; PDB entry 4wbn). Of the 20 crystals screened, only three yielded data suitable for merging based on statistical indicators such as  $R_{\text{meas}}$  and  $\langle I/\sigma(I) \rangle$  (Karplus & Diederichs, 2015). The three crystals diffracted to 2.3 Å resolution. From each crystal, data were collected at two different positions that were well separated. At the first position, a single high-dose 360°  $\omega$  scan, estimated at  $\sim 4$  MGy,

was collected (Table 1 and Supplementary Data; T<sub>2</sub>R-TTL high dose, low multiplicity, single orientation). At the second position, eight low-dose 720°  $\omega$  scans, corresponding to the same accumulated dose of  $\sim 4$  MGy (0.25 MGy per 360°  $\omega$  scan), were collected: seven at different  $\chi$  increments from 0 to 30° at constant  $\varphi$  and an additional  $\omega$  scan at  $\chi = 10$  or 15° and  $\varphi = 90^\circ$  (Table 1 and Supplementary Data; T<sub>2</sub>R-TTL low dose, high multiplicity, multi-orientation).

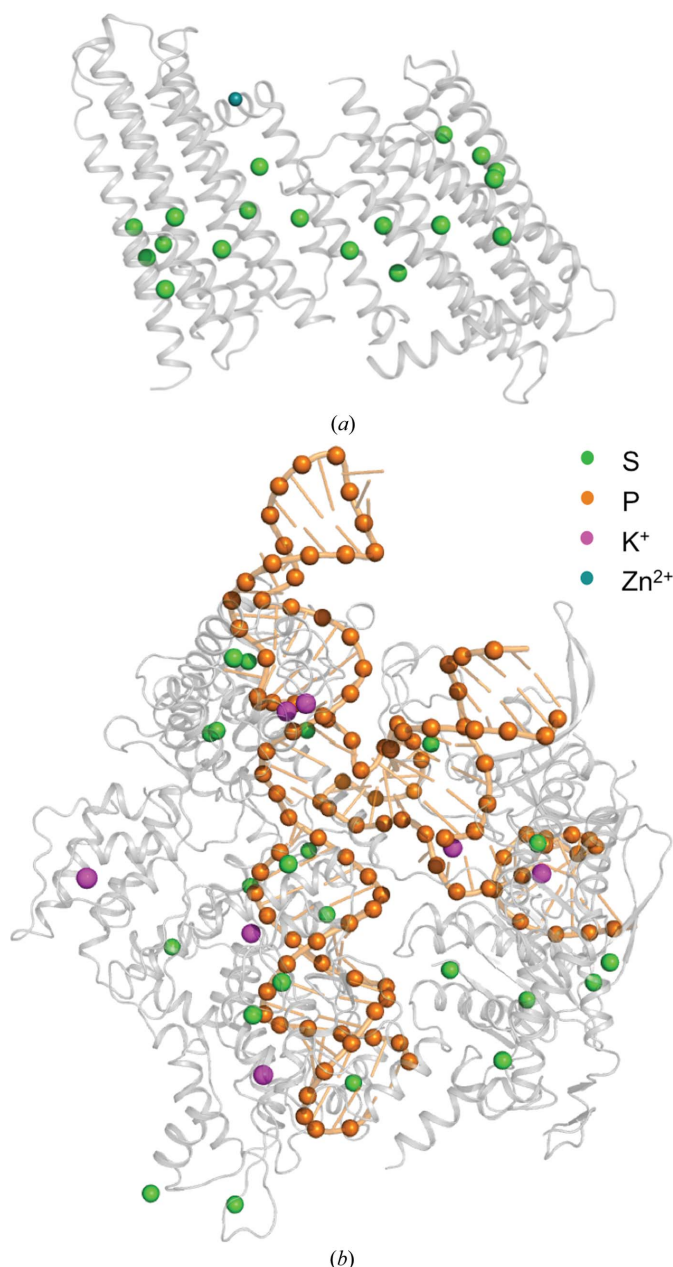
## 2.2. X-ray fluorescence measurements

X-ray fluorescence (XRF) analysis was performed on the PHOENIX beamline at the Swiss Light Source at an incident radiation energy of 4.1 keV, which is above the Ca K edge (4.038 keV) and not currently accessible at most MX beamlines. The crystal was attached directly to carbon tape and loaded into a  $10^{-4}$  Pa vacuum chamber. The energy-dispersive X-ray fluorescence spectrum was recorded at room temperature using a single-element solid-state detector (from Roentec) with 180 eV energy resolution. The X-ray beam measuring  $100 \times 100 \mu\text{m}$  was directed at the center of the crystal. There was no indication of beam damage over the 120 s collection time. The XRF spectrum was fitted using *PyMca* (Solé *et al.*, 2007) to extract the elemental composition of the probed crystal volume. The software provides the elemental composition in terms of single peaks corresponding to different atomic energy-level transitions using Gaussian shape emission lines and energy-dependent photoelectric cross-sections.

## 3. Results and discussion

### 3.1. Native SAD phasing

**3.1.1. DgkA.** The integral membrane diacylglycerol kinase DgkA has two trimers in the asymmetric unit representing  $6 \times 130$  residues with a total molecular weight of 84 kDa. The substructure has  $6 \times 2$  methionines,  $6 \times 1$  cysteines and a zinc ion (Fig. 1a), corresponding to a theoretical Bijvoet ratio of 1.1% at 6 keV. Previous SAD and MAD attempts at the Zn absorption edge were unsuccessful. The structure was originally solved by Se-SAD phasing at the Se peak using crystals diffracting to 2.05 Å resolution (Li *et al.*, 2013). The crystals used in the current study only diffracted to 2.6 Å resolution (see §2.1.1). Data were collected both with and without multiple crystal orientations (Table 1). Substructure identification using *SHELXD* was successful with data cut at 4 Å resolution and with  $E$  values above 1.5; 14 sites out of 19 were identified, 11 of them correctly as determined using *SITCOM* with reference PDB entry 3ze3 (Li *et al.*, 2013). The single correct solution obtained after 1000 *SHELXD* attempts had  $CC_{\text{all}}$  and  $CC_{\text{weak}}$  values of 26.2 and 12.0, respectively (Fig. 2a). The correct hand was identified by density modification in *SHELXE*. Auto-tracing with a search for  $\alpha$ -helices resulted in 364 residues (out of 780) built after ten cycles (Fig. 2b), showing a good cross-correlation to the data of 43.95%. The resulting electron-density map (Fig. 3a) showed a model-map



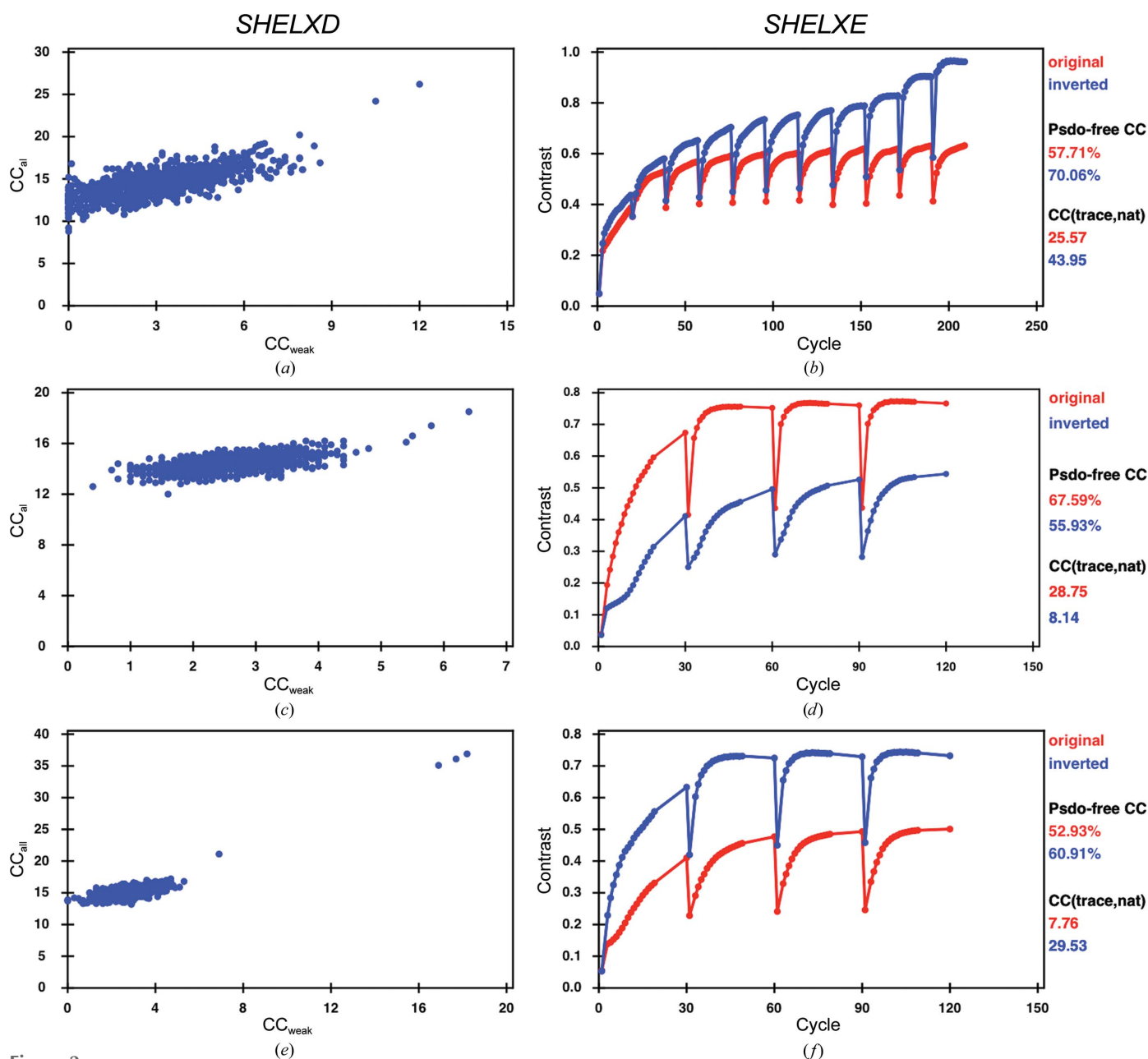
**Figure 1**  
Native SAD structures. (a) DgkA and (b) Cas9-RNA-DNA. The proteins are shown as grey cartoons and the nucleic acid backbones are coloured orange. Anomalous scatterers are depicted as coloured spheres.

cross-correlation of the *SHELXE* map to the final model of 66.8%.

Analysis of the substructure data exemplifies the weakness of the anomalous data, mainly owing to the disordered Met66 and Cys41 in chains *E* and *F*. The experimental Bijvoet ratio of 1% for this sample was therefore lower than the expected theoretical value of 1.1%.

**3.1.2. Cas9–RNA–DNA.** The CRISPR-associated protein Cas9 is an RNA-guided endonuclease that has been repurposed for genome editing and gene-expression control. The Cas9–RNA–DNA complex, which crystallizes in the low-symmetry space group *C*2, was initially solved with SAD

phases obtained from selenomethionine and iridium derivatives (Anders *et al.*, 2014). For native SAD phasing experiments, we used crystals of the Cas9–sgRNA–target DNA complex in which the complementary DNA strand contains mismatches to the guide RNA at positions 1–3 (see §2.1.2). The asymmetric unit contains 1371 amino acids, 83 RNA nucleotides and 39 DNA nucleotides (total molecular weight ~200 kDa), with 24 S atoms and 120 P atoms (Fig. 1*b*), which comprises the largest substructure solved with native SAD phasing to date. In this case, data were collected from three native crystals diffracting to ~2.2 Å resolution in multiple orientations at a wavelength of 2.066 Å and the data were



**Figure 2** *SHELX* output. (a) *SHELXD* substructure-determination output for DgkA for 1000 trials. (b) *SHELXE* output for DgkA after ten cycles of auto-tracing. (c) *SHELXD* substructure-determination output for Cas9 for 1000 trials. (d) *SHELXE* output for Cas9 after three cycles of auto-tracing. (e) *SHELXD* substructure determination for T<sub>2</sub>R-TTL with 1000 trials. (f) Three cycles of chain tracing with phasing and density modification in *SHELXE* for T<sub>2</sub>R-TTL.

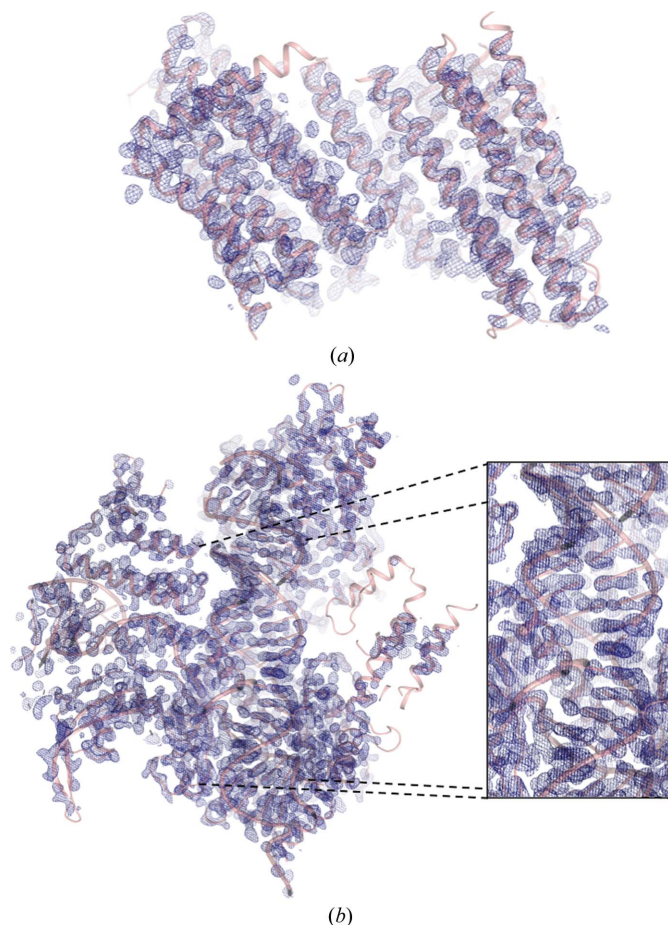
merged in an initial attempt at locating sulfur and phosphorus positions from the anomalous peaks (Table 1). The substructure solution could be obtained with 1000 *SHELXD* trials using a high-resolution cutoff of 2.6 Å in a search for 65 sites, excluding data with *E* values below 1.2 (Fig. 2c). The program initially identified 45 correct sites out of a total of 150 as determined using *SITCOM* with reference PDB entry 4un5 (Anders *et al.*, 2014). *SHELXE* substructure refinement completed the substructure to 114 sites, resulting in a very clear hand separation (Fig. 2d). Three cycles of chain tracing with a search for  $\alpha$ -helices resulted in a readily interpretable map (Fig. 3b) building a C $\alpha$  chain of 1060 residues (out of 1372) after three cycles (Fig. 2b) with a cross-correlation of 28.75% to the data. The model–map cross-correlation of the *SHELXE* map to the final model was 68.0%.

Substructure determination was only successful using data merged from three crystals collected in multiple orientations. The increase in the anomalous peak height resulting from the merging of new data sets (coming from either new orientations of the same crystal or from additional crystals) is shown in Fig. 4. We limited the accumulated dose on each crystal (or crystal location) to a conservative  $\sim 2$  MGy in order to minimize the effects of radiation damage. Additional dose brought little additional information; in the case of crystal 1, we

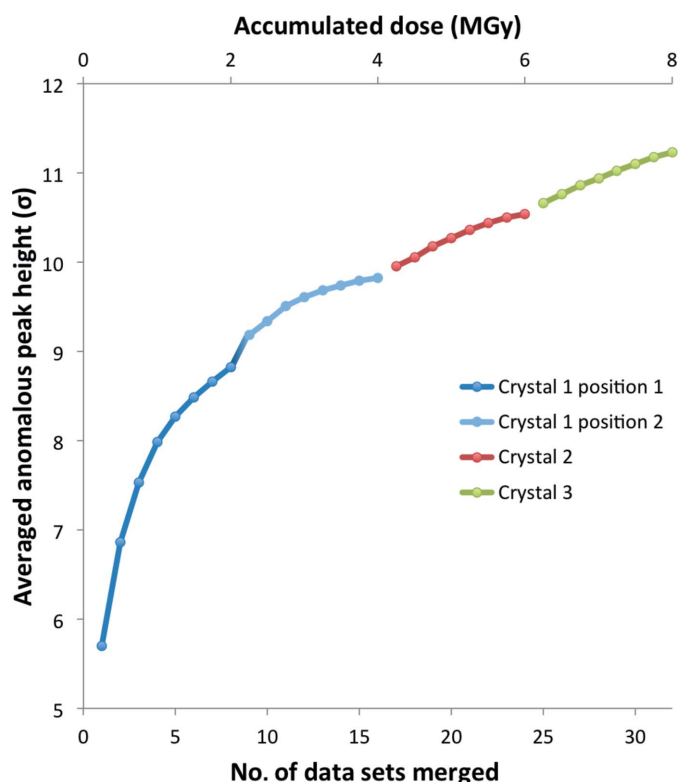
observed that the gain in anomalous signal was small beyond 2 MGy and plateaued at 4 MGy (Supplementary Fig. S2).

### 3.2. X-ray fluorescence analysis

Assigning certain atoms, especially light monoatomic ions, in a crystal structure can be difficult. An added benefit of data collection at low energy is that the anomalous scattering from some of these elements can aid in their correct assignment (Mueller-Dieckmann *et al.*, 2007). In the case of the Cas9–sgRNA–DNA complex, certain ions originally modelled as Mg<sup>2+</sup> (PDB entry 4un5) displayed strong anomalous signal ( $>10\sigma$ ), suggesting alternative atom assignments. However, it was not possible to unambiguously identify the corresponding element from the anomalous peak heights or from an examination of the chemical architecture of the site. One of the crystals was subsequently used to collect a fluorescence spectrum at an incident energy of 4.1 keV (see §2.2). The spectrum confirmed the presence of K<sup>+</sup> (Fig. 5), which was likely to arise from the crystallization condition (the buffer solution contained 250 mM KCl and 300 mM KSCN). We therefore reassigned six Mg<sup>2+</sup> ions as K<sup>+</sup> ions and found an additional one. Subsequent refinement of the anomalous scattering contribution  $f''$ , as implemented in *phenix.refine* (Echols *et al.*, 2014), gave  $f''(\text{K}^+)$  values between 1.58 and 1.72 e, in agreement with the theoretical value of 1.77 e.



**Figure 3**  
Experimental maps after *SHELXE* phasing and density modification, contoured at  $2\sigma$ . (a) DgkA map. (b) Cas9–RNA–DNA map.



**Figure 4**  
Averaged anomalous peak height of the first 70 anomalous peaks for Cas9–RNA–DNA merged data sets. A total of  $3 \times 360^\circ$  data sets were merged from three different crystals. Data from the first crystal were collected at two well separated positions (Supplementary Fig. S1b, crystal 1). The X-ray dose (upper *x* axis) accumulated at each position was around 2 MGy.



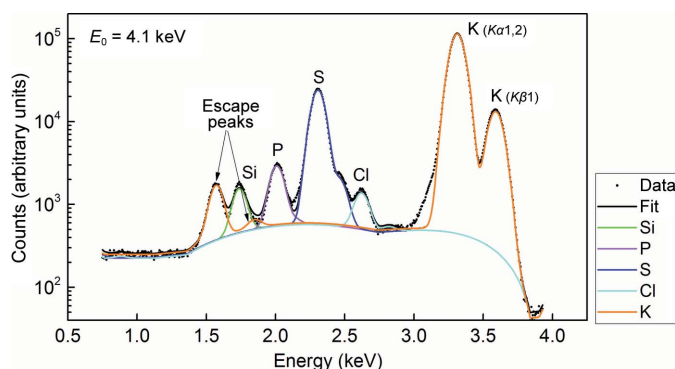
### 3.3. Optimal data-collection strategy for challenging native SAD phasing: T<sub>2</sub>R-TTL as a test case

T<sub>2</sub>R-TTL is a multiprotein/multiligand complex composed of tubulin (T), the microtubule destabilizing- and stathmin-like protein RB3 and the tubulin-modifying enzyme tubulin tyrosine ligase (TTL; Prota, Magiera *et al.*, 2013). The asymmetric unit, composed of two  $\alpha$ -tubulin, two  $\beta$ -tubulin, one RB3, one TTL, two GTP, two GDP and one AMPPNP molecules, consists of 2317 amino acids (total molecular weight of ~266 kDa) and has a substructure with 118 S, 13 P, three Ca and two Cl sites. Following our previous successful structure solution of T<sub>2</sub>R-TTL by native SAD phasing using a single crystal, multi-orientation data-collection strategy (Weinert *et al.*, 2015), we decided to use this system to evaluate the benefits of the combined multi-orientation, multi-crystal strategy. T<sub>2</sub>R-TTL is an excellent example of a challenging test case; it has a large asymmetric unit and a large number of scatterers and, as shown below, it forms crystals of variable quality.

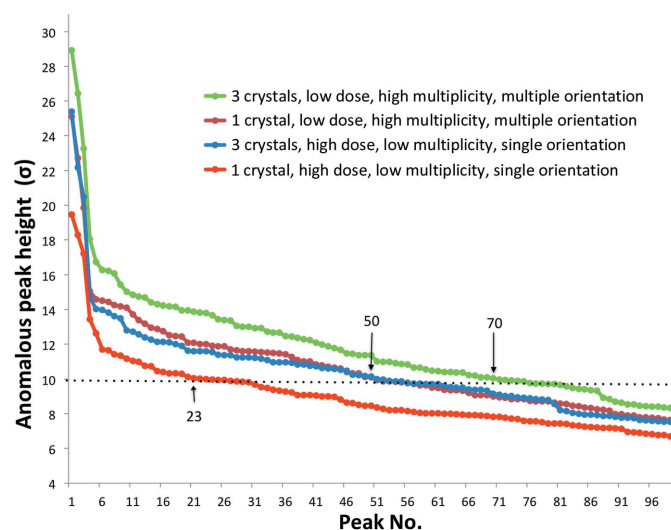
Using multiple native SAD data sets measured from a single T<sub>2</sub>R-TTL crystal (see §2.1.3 and Supplementary Data), it was not possible to determine the substructure. This is likely because the quality of the crystals used in the current work was less than that observed in the original study (Weinert *et al.*, 2015). A substructure solution with *SHELXD* could only be obtained upon merging data sets from three different crystals each measured in eight different orientations (Table 1 and Supplementary Data) in a search for 75 sites using a resolution cutoff of 3.3 Å and with *E* values above 1.5 (Fig. 2e). The successful *SHELXD* substructure solution had a CC<sub>all</sub> and a CC<sub>weak</sub> of 36.9 and 18.2, respectively. The number of correct sites, as determined using *SITCOM* with reference PDB entry 4wnb, was 86 and this increased to 111 using substructure refinement in *SHELXE*. Density modification resulted in a clear separation of hands (Fig. 2f). Three cycles of chain tracing with a search for  $\alpha$ -helices resulted in a readily

interpretable map, building a C <sub>$\alpha$</sub>  chain of 1669 residues (out of 2319) after three cycles with a cross-correlation of 28.75% to the data. The model–map cross-correlation of the *SHELXE* map to the final model was 78.5%.

The merit of combining low-dose data from multiple crystals in multiple orientations, which was readily observed in the merging statistics (see Supplementary Data), was assessed by comparing the average anomalous peak heights for four data-collection strategies: (i) high dose, single orientation from a single crystal; (ii) low dose, multiple orientation from a single crystal; (iii) merging high-dose, single-orientation data from three crystals; and (iv) merging low-dose, multiple-orientation data from three crystals (Fig. 6). A single-orientation scan at high dose gave only 23 anomalous peaks above 10 $\sigma$ , which was insufficient for successful substructure solution. Despite a threefold higher accumulated dose, merging high-dose data from three different crystals resulted in 50 anomalous sites above 10 $\sigma$ , similar to merging low-dose, multiple-orientation data from a single crystal. However, this was still insufficient for successful substructure solution. The substructure search was only successful when low-dose, multiple-orientation data sets from three crystals were merged, yielding 70 anomalous sites above 10 $\sigma$ . This is consistent with our previous study, in which substructure solution and hence native SAD phasing were successful by merging low-dose, multiple-orientation scans from a single crystal, giving 80 anomalous sites above 10 $\sigma$  (Weinert *et al.*, 2015). This is also consistent with a study showing that substructure searches tend to be successful when mean peak heights in the anomalous difference map are above



**Figure 5**  
Log scale of the X-ray fluorescence spectrum of a Cas9–RNA–DNA crystal recorded at an incident energy of 4.1 keV. The measured spectrum is plotted as black dots and the fit as black (total) and coloured (for different elements) lines. The main contributions (orange) originate from the potassium K $\alpha$ 1,2 (*K*–L<sub>III</sub> and *K*–L<sub>II</sub>) and K $\beta$ 1 (*K*–M<sub>III</sub>) characteristic lines. The smaller (orange) peaks between 1.5 and 2 keV represent the K–Si escape peaks and are an artifact of the measurement (Papp & Campbell, 2001). The Si peak (green line) originates from the crystal holder.



**Figure 6**  
Comparison of anomalous peak heights for various data-collection strategies. The anomalous peak heights for the first 100 peaks are shown for four data-collection strategies for T<sub>2</sub>R-TTL data sets: (i) high dose, low multiplicity, single orientation at a single crystal location (red), (ii) low dose, high multiplicity, multiple orientation at a single crystal location (brick), (iii) merging high-dose, single-orientation data from three crystals (blue) and (iv) merging low-dose, multiple-orientation data from three crystals (green). For each crystal, data were collected with the same dose at two well separated positions. The dotted line marks the 10 $\sigma$  threshold. The arrows point to the last peak above 10 $\sigma$  and show the number of sites above this threshold.



a certain threshold (Bunkóczi *et al.*, 2015); for T<sub>2</sub>R-TTL crystals, this threshold is around 10 $\sigma$  (dashed line in Fig. 6).

Taken together, there is a clear benefit in combining the two data-collection strategies: (i) a low-dose, multiple-orientation approach efficiently extracts as much anomalous data as possible while minimizing radiation damage and (ii) if the anomalous signal from a single crystal sample is not sufficient for successful phasing, data sets from multiple statistically equivalent crystals can be merged. Combining the two approaches means that fewer crystal samples are required. In addition, for the same accumulated dose, there is no difference in diffraction resolution between data sets collected using a high-dose, single-orientation strategy and merged data sets collected using a low-dose, multiple-orientation strategy (Table 1 and Supplementary Data).

#### 4. Conclusion and outlook

Native SAD could be considered the ‘ideal’ *de novo* phasing method for macromolecules because it relies solely on information from atoms that are naturally present in the sample. The method is readily applicable to challenging cases. Accordingly, there is every reason to try native SAD first with macromolecules that contain a sufficient number of naturally occurring anomalous scatterers. Here, we show that multi-crystal averaging of data sets used in conjunction with a low-dose, multi-orientation strategy is beneficial in difficult cases where data from one crystal are insufficient. Further, low-dose data collection in multiple orientations means that fewer crystal specimens are needed to provide the required data accuracy.

That said, this combined approach is still in its infancy and there is room for improvement in order to make native SAD more broadly applicable and more generally accepted, especially in cases of low resolution, which are currently dominated by heavy-atom phasing methods. In addition, collecting multiple 360° scans on multiple crystals, as detailed here, takes more time than single-orientation high-dose scans and may be unappealing in situations where synchrotron beamtime is limited. Nevertheless, advances in data-acquisition hardware, in particular detector instrumentation, will rapidly improve the situation. Furthermore, the method described here lends itself to automation.

As the frontiers of crystallography continue to expand, fundamental techniques for solving the phase problem will continue to evolve. In particular, new methods such as serial crystallography have presented new challenges that require the development of tools for obtaining the most accurate measurements. Data collection at energies below 6 keV should boost the anomalous signal, but presents technical challenges, especially air absorption and scattering, sample absorption and detector geometry. These are currently being investigated at operational beamlines such as BL-1A at the Photon Factory (see Rose *et al.*, 2015), as well as upcoming beamlines such as I23 at Diamond Light Source (Allan *et al.*, 2015) and LAX at NSLS-II (Hendrickson, 2014).

#### Acknowledgements

This work was supported in part by funding from Science Foundation Ireland (12/IA/1255 to MC), by a grant from the Swiss National Science Foundation (310030B\_138659 to MOS) and a European Research Council Starting Grant (ANTIVIRNA, Grant no. 337284 to MJ).

#### References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
- Akey, D. L., Brown, W. C., Konwerski, J. R., Ogata, C. M. & Smith, J. L. (2014). *Acta Cryst.* **D70**, 2719–2729.
- Allan, D. R., Collins, S. P., Evans, G., Hall, D., McAuley, K., Owen, R. L., Sorensen, T., Tang, C. C., von Delft, F., Wagner, A. & Wilhelm, H. (2015). *Eur. Phys. J. Plus*, **130**, 56.
- Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. (2014). *Nature (London)*, **513**, 569–573.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Brockhauser, S., Ravelli, R. B. G. & McCarthy, A. A. (2013). *Acta Cryst.* **D69**, 1241–1251.
- Bunkóczi, G., McCoy, A. J., Echols, N., Grosse-Kunstleve, R. W., Adams, P. D., Holton, J. M., Read, R. J. & Terwilliger, T. C. (2015). *Nature Methods*, **12**, 127–130.
- Dall’Antonia, F. & Schneider, T. R. (2006). *J. Appl. Cryst.* **39**, 618–619.
- Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 688–696.
- DeLano, W. L. (2002). *PyMOL*. <http://www.pymol.org>.
- Echols, N., Morshed, N., Afonine, P. V., McCoy, A. J., Miller, M. D., Read, R. J., Richardson, J. S., Terwilliger, T. C. & Adams, P. D. (2014). *Acta Cryst.* **D70**, 1104–1114.
- El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D. R., Harlos, K., Grimes, J. M. & Stuart, D. I. (2014). *Acta Cryst.* **D70**, 2197–2203.
- Finke, A. D., Panepucci, E., Vornrhein, C., Wang, M., Bricogne, G. & Olieric, V. (2016). *Methods Mol. Biol.* **1320**, 175–191.
- Hendrickson, W. A. (2014). *Q. Rev. Biophys.* **47**, 49–93.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Henrich, B., Bergamaschi, A., Broennimann, C., Dinapoli, R., Eikenberry, E. F., Johnson, I., Kobas, M., Kraft, P., Mozzanica, A. & Schmitt, B. (2009). *Nucl. Instrum. Methods Phys. Res. A*, **607**, 247–249.
- Huang, C.-Y., Olieric, V., Ma, P., Panepucci, E., Diederichs, K., Wang, M. & Caffrey, M. (2015). *Acta Cryst.* **D71**, 1238–1256.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Karplus, P. A. & Diederichs, K. (2015). *Curr. Opin. Struct. Biol.* **34**, 60–68.
- Klinke, S., Foos, N., Rinaldi, J. J., Paris, G., Goldbaum, F. A., Legrand, P., Guimarães, B. G. & Thompson, A. (2015). *Acta Cryst.* **D71**, 1433–1443.
- Li, D., Lyons, J. A., Pye, V. E., Vogeley, L., Aragão, D., Kenyon, C. P., Shah, S. T. A., Doherty, C., Aherne, M. & Caffrey, M. (2013). *Nature (London)*, **497**, 521–524.
- Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancina, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.
- Liu, Q., Guo, Y., Chang, Y., Cai, Z., Assur, Z., Mancina, F., Greene, M. I. & Hendrickson, W. A. (2014). *Acta Cryst.* **D70**, 2544–2557.
- Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Cryst.* **D69**, 1314–1332.
- Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. & Weiss, M. S. (2007). *Acta Cryst.* **D63**, 366–380.

- Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* **D61**, 1263–1272.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Papp, T. & Campbell, J. L. (2001). *X-ray Spectrom.* **30**, 77–82.
- Prota, A. E., Bargsten, K., Zurwerra, D., Field, J. J., Díaz, J. F., Altmann, K.-H. & Steinmetz, M. O. (2013). *Science*, **339**, 587–590.
- Prota, A. E., Magiera, M. M., Kuijpers, M., Bargsten, K., Frey, D., Wieser, M., Jaussi, R., Hoogenraad, C. C., Kammerer, R. A., Janke, C. & Steinmetz, M. O. (2013). *J. Cell Biol.* **200**, 259–270.
- Rose, J. P., Wang, B.-C. & Weiss, M. S. (2015). *IUCrJ*, **2**, 431–440.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Solé, V. A., Papillon, E., Cotte, M., Walter, P. & Susini, J. (2007). *At. Spectrosc.* **62**, 63–68.
- Thorn, A. & Sheldrick, G. M. (2011). *J. Appl. Cryst.* **44**, 1285–1287.
- Waltersperger, S., Olieric, V., Pradervand, C., Gletting, W., Salathe, M., Fuchs, M. R., Curtin, A., Wang, X., Ebner, S., Panepucci, E., Weinert, T., Schulze-Bries, C. & Wang, M. (2015). *J. Synchrotron Rad.* **22**, 895–900.
- Weinert, T. *et al.* (2015). *Nature Methods*, **12**, 131–133.
- Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *J. Appl. Cryst.* **46**, 1225–1230.